Structural and functional analysis of an alternatively spliced chicken TK messenger RNA

Gary F. Merrill and Frank D. Tufaro*

Department of Biochemistry and Biophysics, Oregon State University, Corvallis, OR 97331 and
*Department of Embryology, Carnegie Institute of Washington, 115 West University Parkway,
Baltimore, MD 21210, USA

Received 12 May 1986; Revised and Accepted 16 July 1986

ABSTRACT

The nucleotide sequence of independent cDNA clones revealed that TK transcripts can undergo alternative splicing within the 3' nontranslated portion of the message. The most abundant mRNA is 2.1 kb in length and is colinear with the underlying genomic sequence for at least the first 1052 bases of its 3' nontranslated region. A less abundant mRNA, 1.3 kb in length, differs from the larger mRNA in that an 863 base intron is spliced from the 3' nontranslated region. The possibility that removal of the alternatively spliced intron allows the small mRNA to persist in postreplicative cells was investigated in two ways. First, the levels of large and small TK mRNA in tissues expressing differential growth rates was determined. Second, the pattern of TK enzyme expression during differentiation was analyzed in muscle cells transformed with recombinant TK genes lacking the 3' nontranslated intron. Both lines of experimentation indicated that the small TK mRNA was as dependent as the large TK mRNA on the replicative state of the cell.

INTRODUCTION

Thymidine kinase (TK) is representative of a class of enzymes engaged in DNA precursor biosynthesis and preferentially expressed in replicating cells. RNA blot hybridization analyses have shown that changes in the growth state of cultured cells are accompanied by changes in the cellular level of TK mRNA (1,2,3). On the basis of nuclear run-off determinations, Groudine and Casimir (4) concluded that steady state levels of TK mRNA are governed primarily by a post-transcriptional mechanism. A post-transcriptional level of regulation is consistent with results of in vitro mutagenesis/cell transformation experiments showing that replication-dependent expression of TK activity is not governed by the transcriptional promoter of the cellular TK gene (2). Post-transcriptional regulation has also been invoked to explain replication-dependent expression of DHFR (5,6), although this view has been contested (7). If expression of replicative enzymes is controlled at a post-transcriptional level, the cis-acting regulatory information must be present on the RNA precursor or mature mRNA. An understanding of the genesis,

structure and fate of the RNAs encoding replicative enzymes is therefore basic to understanding how this class of gene products is regulated.

The predominant messenger RNA encoding chicken thymidine kinase (TK) is 2,100 residues in length (9,4). Based on L cell transformation studies the chicken TK gene was assumed to reside entirely within a 3.0 kb Hind 3 restriction fragment (9,10). Indeed, the sequence of a TK cDNA clone isolated from chicken testes RNA established that the entire protein coding region of the TK genes lies within this 3.0 kb genomic fragment. However, 73 nucleotides downstream from the translation stop codon the sequence of the testes-derived TK cDNA diverged from the known genomic sequence at a donor splice site consensus sequence (11). This result was unexpected, as nuclease protection experiments had indicated that TK mRNA was colinear with the underlying genomic sequence for at least 700 bases downstream from the translation stop codon (9). It thus seemed likely that TK transcripts were capable of undergoing alternative splicing within the 3' nontranslated portion of the message.

In the present report we show that the TK transcription unit extends further in a 3' direction than previously believed. We further show that in addition to the major 2.1 kb TK mRNA, a second, less abundant transcript is generated through removal of an 863 base intron from the 3' nontranslated portion of the larger message. To investigate the possibility that the smaller message escapes replication dependent expression, the occurrence of the alternatively spliced TK mRNA was determined in several embryonic and post-embryonic chicken tissues exhibiting different growth states. In addition the pattern of TK regulation was determined in mouse muscle cells transformed with a recombinant TK gene from which the alternatively-spliced intron was removed. Our results indicate that alternative splicing does not result in escape from growth-dependent expression.

MATERIALS AND METHODS

Preparation of and Northern Blot Analysis of RNA

The source of RNA for Northern blot analysis was white leghorn chickens, obtained from the Department of Poultry Science, Oregon State University. Animals were sacrificed by decapitation at 11 days post-fertilization or two days after hatching (24 days post-fertilization). Testes RNA was obtained from a 4 month adult rooster. RNA was prepared by a modification of the method of Glisin et al. (12). Immediately after removal from each animal, tissues were homogenized, using a Kimatica GmbH at maximum setting, in at

least five volumes of 4 M guanidinium isothiocyanate, 5 mM sodium citrate (pH 7.0), 0.1 M β -mercaptoethanol, 0.5% Sarkosyl. Homogenates were sonicated for one minute, using a Branson Sonifier at maximum setting, to shear DNA and reduce viscosity. Cesium chloride was added to 2.3 M, and the homogenate was layered onto 0.3 volumes of a 5.7 M CsCl, 0.1 M EDTA cushion in a polyallomer ultracentrifuge tube. Following centrifugation for 18 hours at 85,000 x g, the clear lenticular RNA pellet was rinsed once with 70% EtOH, resuspended in 0.4 ml 1X TES (10 mM Tris, pH 7.4, 5 mM EDTA, 1% SDS), phenol/CHCl₃ extracted, CHCl₃ extracted and ethanol precipitated. The ethanol precipitate was rinsed once with 70% EtOH and resuspended in a minimum volume of H₂O. Polyadenylated RNA was prepared as described by Aviv and Leder (13) using oligo d(pT) cellulose (12-18 residues) (Collaborative Research). Yield of total and polyadenylated RNA was determined by A₂₆₀.

Electrophoresis of RNA on 1.1% agarose, 2.2 M formaldehyde gels was as described by Lehrach et al. (14) except that 20 mM morpholinopropanesulfonate (MOPS), pH 7.0 was used as buffer in place of sodium phosphate. After electrophoresis, gels were rinsed several times with water, stained with EtBr, photographed, soaked for 30 minutes in 20X SSC (3 M NaCl, 0.3 M sodium citrate, pH 7.0) and blotted overnight onto BA85 nitrocellulose (Schleicher and Schuell) using 20X SSC. Blots were air-dried for 30 minutes and baked at 80° for 4 hours. Overnight prehybridization was at 65° in 5X SSC, 25 mM sodium phosphate (pH 6.5), 50% formamide, .025% denatured salmon sperm DNA, 0.02% bovine serum albumin, 0.02% Ficoll 400 and 0.02% polyvinylpyrrolidone. Hybridizations was for 2 days at 65° in prehybridization buffer (25 μ l per cm²) supplemented with 10% (w/v) dextran sulfate (Pharmacia) and 10⁶ cpm/ml of ³²P-labeled synthetic RNA probe. Blots were washed twice for 5 minutes at 20°, then twice for 30 minutes at 65° in 50 mM NaCl, 20 mM sodium phosphate (pH 6.5), 1 mM EDTA and 0.1% SDS. Non-specific hybridization of probe to ribosomal RNA was frequently observed. To eliminate this background, blots were incubated for 15 minutes at 20° in 2X SSC containing 0.1 μ g/ml RNase A (Boehringer), followed by one 15 minute rinse at 65° in 2X SSC. Blots were exposed for one to ten days to XAR X-ray film (Kodak) using a Cronex Hi Plus intensifying screen (Dupont).

Labeled RNA probes were synthesized using materials and procedures supplied with the Riboprobe System (Promega). Unlabeled rUTP was omitted from the transcription reaction and ³²P-rUTP (800 Ci/mmol; New England Nuclear) was present at 5 μ Ci/ μ l. The template used to make probe complementary to the protein-coding region ("coding probe") was made by

cloning a 1568 Eco R1/Bam H1 fragment from $\Delta 3'734$, a 3' deletion mutant of the chicken TK gene (11), into SP64 and linearizing the template at an Sst 1 site within the 3rd exon. The template used to make probe complementary to sequences exclusive to the 3', alternatively spliced, non-translated intron was made by cloning a 1550 bp Eco R1/Hind 3 "cDNA fragment" into SP64 and linearizing the template at a Pvu 2 site 175 bp downstream from the translation stop codon. Greater than 90% of the label was incorporated during all transcription reactions.

Construction of cDNA Libraries

Construction of a testes cDNA library in λ gt10 was described elsewhere (11). The bursa lymphoma cDNA library was constructed by Steven McKnight using the same procedure. The source of the bursa lymphoma cell RNA was RP9C, a rapidly-growing B-cell line transformed by the RAV-1 virus (4). Two TK-positive plaques were identified and shown to contain 1.4 and 1.7 kb inserts. Inserts were subcloned into pBR322 and regions complementary to chicken TK were sequenced by the chemical method (15). Both bursa lymphoma cDNA clones contained non-relevant sequences at one or both termini; these were presumably artifacts of ligation steps in the cloning procedure.

Genomic Sequence Analysis

A 12.5 kb genomic phage clone, λ ch-tk1, (16) served as source of a 1.47 kb Hind 3/synthetic Eco R1 fragment flanking the 3' end of the previously characterized 3.0 kb Hind 3 fragment. Phage DNA was cut with Eco R1, filled-in with AMV reverse transcriptase, and Bam H1 linkers were attached. After digestion with Hind 3 and Bam H1 the 1.47 kb Hind 3/Bam H1 fragment was inserted into Hind 3/Bam H1 cleaved M13mp19. After digestion of the TK-containing M13 DNA with Bam H1 and Kpn 1, a nested series of deletion mutants was generated by preferential resection of 5' overhangs by Exonuclease III, followed by treatment with S1 nuclease and recircularization (17). Nucleotide sequence was then determined by the dideoxynucleotide termination method, using the M13 17-base sequencing primer (18).

Cell Growth and Transformation

Mouse L cells, a sub-strain of the TK⁻ line derived by Kit et al. (19), were grown in Dulbeccos Modified MEM (GIBCO), supplemented with 10% calf serum (Hyclone) and 1% antibiotics (GIBCO). Cells were transformed with 1.0 μ g of TK-containing plasmid by the calcium-phosphate precipitation method of Graham and van der Eb (20), as modified by Corsaro and Pearson (21). Selection was in growth medium supplemented with 10^{-4} M hypoxanthine, 4×10^{-7} M aminopterin and 10^{-5} M thymidine. About 500 transformants were

obtained per 5×10^5 cells exposed to DNA. RNA was harvested from L cell transformants when cell density reached about 2×10^7 cells per 10 cm dish.

TK⁻A1e mouse muscle cells, a subline of the TK⁻A line described previously (22), were grown on gelatin-coated culture dishes in Ham's F10, supplemented with 0.8 mM calcium chloride, 15% horse serum, 3% chicken embryo extract and 1% antibiotics solution (10 U of penicillin G and 0.5 mg of streptomycin sulfate per ml). Myoblasts were transformed as described above for mouse L cells except that: plasmids were linearized with Sal I prior to transfection; cells were exposed to DNA for 6 hours rather than 24 hours; cells were passaged into three 10 cm dishes containing HAT at 24 hours after DNA addition. To induce differentiation muscle cells were rinsed twice with Ham's F10 and incubated in defined mitogen-depleted medium (Ham's F10 supplemented with 0.8 mM calcium chloride, 10^{-6} M insulin and 10^{-7} M dexamethasone. In the absence of FGF-like mitogens, mouse myoblasts irreversibly withdraw from the cell cycle and commit to terminal differentiation within a single cell generation time (23).

RESULTS

Analysis of Genomic and cDNA Sequences in the 3' Region of the TK Gene

Transcripts encoding replicative enzymes are rare. For example, even in exponentially growing chicken cell populations, there are fewer than five TK mRNA molecules per cell (Gross and Merrill, in preparation). The low copy number of TK transcripts has probably contributed to the poor representation of TK-containing clones in cDNA libraries. Thus far, three chicken TK-containing cDNA clones have been isolated and characterized. All three cDNA clones extended in a 3' direction beyond the previously sequenced 3.0 kb Hind III fragment. Evidently, although it efficiently rescues TK⁻ cells and contains the entire protein-coding region, the 3.0 kb Hind III fragment lacks the 3' terminal portion of the TK gene. Therefore, to obtain a more complete picture of the TK transcription unit, an adjoining 1.3 H3/R1 genomic fragment was subcloned from phage clone λ chtk-1 (16) and sequenced. Genomic and cDNA sequence information in the 3' region of the TK gene is shown in Figure 1. Two TK-containing cDNA clones, isolated from a library derived from bursa lymphoma cell RNA, were colinear with the underlying genomic sequence for at least 1022 nucleotides downstream from the translation stop codon. A third cDNA clone, isolated from a library derived from adult chicken testes RNA, differed from the bursa lymphoma-derived cDNAs in that it diverged from the underlying genomic sequence at a GT dinucleotide 73 bp

```

bgl 2      uga
AGATCTTGGC TTCTGATTG CTGCGCTCCA GCGTGGAGAG GGATGGATAC AAAGAAAGAA GATAAATGTT GCAGCTCCTG GACTTCAGGT CTGCTTCCC
-----
CCCTTTTACC AAAACTGTCA TATGCCATAG CTCTACCTGC CAAACCAAGC TGCTCTTCTT CTGCTCCAAG TACCAGGAAA GAGGGCAATC AGCTGGACCT
-----
GGTGGCAGGG AGAGCTGAAA CAGAATCTGA CTAAAAGAGA GTTCTTTTCC CAGTACGTGG AAAAGAGGTT TAATTACTGA TCTGCCATA CAGATGTGCT
-----
GGAAAGCCAC CACCCCAATC CAGGCTGTGG GGCACGTACA ACTTGCCATA GCTCTTCTGG TCTGTCTCTT GCTCCTGGCA CATGTCAGA CTTGCCACTG
-----
GTAGCCAGGT CATTGGCCAC CAAACAGCAA TTTTCTCTC ATGTGCTGTT GCTAAATCTT GGGTTTCTTA AGACAGTGTA CAGTAACCTG ACTTCCTAAA
-----
GCCATTTCCT TTTTAATAGA GCGTGAGGAG ATTCCCTTTC CCTTACTGTG GAACAGTGGC ACTACTTCAA CTCACACATT GCAGGACGCT GCCTCTCTGC
-----
AGTCTCTTAA CAATTGTGTA TAGCCAATTA GCTTGTCTGT AAATCTGCTG TAAATACATG GGGTTTTTGT TTTGTTTGT TTTAAATGGA GGACAGGAGA
-----
AGGAAGGCTT GACTTGTGAC CCGGACTCCT AGGCTCATTC TAGCCTTGA AATTAAAAAG GATTTTCTGC ACAGCTTGG TGCAGGCCAC TTTGTAAGGG
-----
TTAAGTGACC TCTGGAGGGT TTGGGGGATC AGGAATTAAC CTGGTATCAG AGTGGCTACC TTCTCTGTGA AGATGGAACG TACCTGGAAA AGGGTCTTTT
-----
CTTTGAGTGG ATGGTTCTCA GCTCTAGTCC TGTGTTCCGC TTCTCTCTCA GCCTTCCAAG AATCCTTGAG CCGTGTGCTT TCCTGTGTGA ACACCTGAAG
-----
AATCCTCTCC GTTGCTTGAT GTAGCTTCTG GAAGGAATTT TACTGCTTGT AATAAAAGTG CTTTTTCTT CATTCAGGA TAACCTTTGA GCTCTGAGCG
-----
GTAGCTCTTG GTTCTGGCTG GGCTTTGGTC AATCCCTCTC GAAATACAGT GCTCTCTCCA ATACTCTCTC CAATACTGT TACCACTTGA TTAGGAGGAG
-----
GCAGCTGCCC AGGTGGTTAC AAGACTGAAC CCTCGCCTTC CCGTGTACC AAACAATTAC AGCTATTGGT GTTCACAAAA GCTCGTCAGG CCTTGGGCTT
-----
TCCCTTtagG GAGGGGGGAC AGGCAGCAGT GCTTATATTC CTACATTAAT AAATAACTTA ATGAAATTAC CTGTGACTTA TCGGAGCGAC TGGTCACAGC
-----
GCACTGATGT GCTCAGGCTG TCCTGAGGAG CAGCAAGCCT TGTAATAATA GGTGCTGGCG TGATGTGTGG GAAACCCCTG GCGTTACCCA ACTTAATGCG
-----
CTTGACGAC ATCCCCCTTT GCCCAGCTG

```

Figure 1: Genomic and cDNA nucleotide sequence in the 3' region of the chicken TK gene. The genomic sequence shown extends from a unique Bgl 2 restriction site 15 bases upstream from the UGA translation stop codon to a non-unique Pvu 2 site located 1509 bases downstream from the stop codon. The genomic sequence between the Bgl 2 and Hind 3 restriction sites was published earlier (11); it is reproduced here to facilitate description of the alternatively spliced 3' intron. The two solid lines below the genomic sequence represent the sequence of bursa lymphoma cDNA clones I (top) and II (bottom). The broken line represents the sequence of a testes cDNA clone, and shows the location of the intron donor and acceptor site of the alternatively spliced 3' non-translated intron. The putative AAUAAA polyadenylation signal is two bases downstream from the 3' terminus of the testes cDNA.

downstream from the translation stop site. The testes cDNA sequence rejoined the genomic sequence at an AG dinucleotide 936 bp downstream from the translation stop site. This result indicated that, at least in certain

cells, an 863 base intron is spliced from the 3' nontranslated portion of the TK mRNA.

The testes-type splicing pattern does not bring an alternative polyadenylation signal into juxtaposition with the TK protein-coding region. After rejoining the genomic sequence, the testes-derived cDNA extends colinearly with the bursa lymphoma-derived cDNAs for the remaining length of each clone. None of the three TK cDNA clones contained a tract of polyadenylate residues. Presumably exonucleolytic activity during the cDNA cloning procedure resulted in loss of the native 3' structure. Of the three cDNAs, the testes-derived clone extended furthest in a 3' direction. Inspection of the genomic sequence immediately downstream from the 3' terminus of the testes cDNA (Figure 1) revealed an AATAAA consensus sequence known to serve a polyadenylation function in several eucaryotic mRNAs (24). RNase protection experiments with a SP6-generated RNA probe spanning this putative polyadenylation signal gave a complex pattern of protected fragments, possibly due to the presence of internal RNase sensitive sites in the RNA/RNA duplex. Positive identification of the 3' terminus of TK mRNA will require isolation and sequencing of cDNA clones containing polyadenylic acid.

Assuming that the AATAAA hexanucleotide denoted in Figure 1 is the signal specifying polyadenylation of both the testes- and bursa lymphoma-type TK mRNAs and that both mRNA species utilize a common transcription initiation site, a model of the TK transcription unit was constructed (Figure 2). Given that the protein coding region of the TK gene is 669 nucleotides in length and is preceded by a nontranslated leader of about 110 residues (11), a message unspliced in the 3' nontranslated region would have a predicted length of about 2100 residues. Removal of an 863 base intron from the 3' nontranslated region would generate a message approximately 1300 residues in length. The presence and length of attached polyadenylate residues would affect these size estimates modestly. To substantiate this model, Northern blots were done to ascertain the sizes of TK mRNAs produced by various chicken tissues and mouse cells transformed with recombinant chicken TK genes.

Northern Blot Analysis of TK Transcript Length

The model presented in Figure 2 predicts that alternative splicing of precursor transcripts from the chicken TK locus can generate two different-sized TK mRNAs, with lengths of 2.1 and 1.3 kb, respectively. It should be possible to document the existence of these two messages by Northern blot analysis. Previous analyses of RNA from chicken thymus and

Figure 2. Schematic diagram of the chicken TK transcription unit. At top is shown a restriction map of the chicken TK locus (16). H, R and Bg represent Hind 3, Eco R1 and Bgl 2 restriction enzyme sites. Open arrows represent extent and polarity of SP6-generated RNAs used to probe Northern blots. The "coding probe" was synthesized from an 1819 bp template extending from a Bam H1 linker 24 bp downstream from the translation stop codon to a Sst 1 site within the 3rd exon of the protein-coding region. The "intron probe" was synthesized from a 593 bp template extending from the right-hand Hind 3 site to a Pvu 2 site 175 bp downstream from the translation stop codon. At center is depicted the splicing pathway used to generate the 2.1 kb TK mRNA; the extent of overlap with both bursa lymphoma cDNA clones is shown. At bottom is depicted the splicing pathway used to generate the 1.3 kb mRNA; the extent of overlap with the testes cDNA clone is shown. Hatched regions denote the protein-coding domain.

bursa tissue, as well as RNA from proliferating cultures of chicken embryo fibroblasts, transformed T-cells and bursa lymphoma cells, all showed a single band at 2.1 kb (4,3). The 1.3 kb mRNA was either extremely rare or absent in these cell- and tissue-types.

In searching for tissues capable of producing the 1.3 kb message, we were particularly interested in the possibility that alternative splicing might represent a mechanism for maintaining a low level of TK mRNA in terminally differentiated cells. Such a mechanism would be evolutionarily advantageous

if post-replicative cells required a small basal level of phosphorylated thymidine for DNA repair or other metabolic processes. In view of evidence that TK mRNA levels are post-transcriptionally regulated (4), alternative splicing might remove a structure responsible for TK mRNA instability in non-replicating cells. If this were the case, levels of the 2.1 kb mRNA would decline as cells withdrew from the cell cycle, but levels of the 1.3 kb mRNA would remain unchanged. This would explain why a cDNA clone representative of the 1.3 kb message was isolated from a library derived from adult tissue, whereas two cDNA clones representative of the 2.1 kb mRNA were isolated from a library derived from proliferating cells in culture.

To identify tissues capable of producing TK transcripts other than the 2.1 kb species and address the question of differential dependence of the two TK mRNA species on cell replication, polyadenylated RNA was prepared from several embryonic and post-embryonic chicken tissues and analyzed by Northern blotting. Blots were hybridized with either the "coding probe" or "intron probe" shown in Figure 2. The coding probe would be expected to hybridize with both the 2.1 and 1.3 kb mRNAs, whereas the intron probe should only hybridize with the 2.1 kb mRNA. To detect a possibly rare 1.3 kb message, blots were purposefully overloaded and given long autoradiographic exposures.

Figure 3A shows a blot hybridized with the protein coding region probe. Embryonic muscle, heart and liver RNA preparations gave a readily detectable signal at 2.1 kb (lanes 1, 2 and 3, respectively). A prominent 2.1 kb signal was also detected in RNA from embryonic brain, kidney, spleen, gizzard, skin and total carcass (data not shown). Of the embryonic tissues surveyed, only heart (lane 2) gave a detectable hybridization signal at a position corresponding to 1.3 kb. In addition to having the electrophoretic mobility expected for the alternatively spliced TK mRNA, the signal at 1.3 kb also exhibited the predicted probe specificity. As shown in Figure 3B, the 1.3 kb signal could not be detected in embryonic heart (lane 2) in blots hybridized with the probe specific for the 3' nontranslated intron ("intron probe" in Figure 2). As expected, the signal at 2.1 kb was detected using both the coding probe and the intron probe. Since it exhibited both the correct size and correct probe specificity, we conclude that the 1.3 kb signal detectable in embryonic heart tissue corresponds to the alternatively spliced mRNA shown in Figure 2.

The results in Figure 3A also address the question of the relative abundance of the two TK mRNA species in rapidly proliferating versus post-mitotic tissues. Proliferative rate, as measured by rate of change in

Figure 3. Northern blot analysis of TK messenger RNA produced by embryonic and post-embryonic chicken tissues. RNA samples were run on formaldehyde gels, blotted onto nitrocellulose and hybridized with either the "coding" probe (A) or "intron" probe (B) shown in Figure 2. Each lane contains 15 μ g of poly A⁺ RNA from: embryonic muscle (1), heart (2) and liver (3); hatching muscle (4), heart (5) and liver (6); adult testes (7). Embryonic and hatching tissues were prepared at 11 and 24 days post-fertilization, respectively. To eliminate non-specific binding to 28S rRNA, hybridized blots were treated with RNase A prior to 10 day autoradiographic exposure. Bars indicate the expected mobilities for RNAs 1.3 and 2.1 kb in length, based on EtBr-stained Hind 3-digested λ phage markers.

DNA content per organ, is several fold lower in hatching muscle, heart and liver than in embryonic muscle, heart and liver (Gross and Merrill, manuscript in preparation). Hatching muscle, heart and liver (lanes 4, 5 and 6, respectively) clearly contained less 2.1 kb mRNA than corresponding embryonic tissues (lane 1, 2 and 3 respectively). This reinforces the idea that levels of TK enzyme and the major TK mRNA species are dependent on the replicative state of the tissue of origin. The reduced intensity of the 1.3 kb signal in hatching heart as compared to embryonic heart (lanes 2 and 5, Figure 3A) indicates that the small TK mRNA is also at least partially dependent on the replicative state of the tissue. The low level of small TK mRNA even in embryonic heart tissue made it difficult to determine whether the magnitude

Figure 4. Northern blot analysis of TK mRNA produced by mouse L cells transformed with TK genes or chicken embryo fibroblasts. Each lane contains 5 μ g of poly A⁺ RNA from proliferating cultures of chicken embryo fibroblasts (lane 4) or L cells transformed with plasmid containing: the 3.0 kb Hind 3 fragment (lane 1); the 4.3 kb Hind 3/Eco R1 fragment (lane 2); a modified TK gene in which the 3' nontranslated intron was removed (lane 3). Blots were hybridized with the "coding probe" shown in Figure 2. Numbers at right indicate the expected mobilities for RNAs 2.1, 1.6 and 1.3 kb in length, based on EtBr-stained phage λ DNA markers.

of the decline in small message at hatching was as great as the decline in large message. To pursue the question of differential sensitivity of the two TK mRNAs to changes in replicative state, the pattern of TK expression in mouse cells transfected with recombinant TK genes was investigated as described below.

Our expectation that the 1.3 kb mRNA would be the prevalent TK mRNA species in adult testes was incorrect. As can be seen in lane 7 of Figure 3A, adult testes was as rich in 2.1 kb message as embryonic tissues. The 1.3 kb mRNA constituted only a small percentage of testes TK mRNA. Persistence of TK mRNA in adult testes tissue could reflect the presence of replicating

spermatogenic cells in this adult tissue. In retrospect, the occurrence of a cDNA clone representative of the 1.3 kb mRNA in the testes library was probably due to its shorter nontranslated region. Since oligo-dT was used to prime first strand synthesis during library construction and since a protein-coding region probe was used to screen the library, the rare 1.3 kb mRNA may be over-represented in the library because reverse transcriptase would only have to travel a short distance before reaching the protein coding region on the small TK mRNA.

Analysis of TK Expression in Mouse Cell Transformants

Northern blot analysis of polyadenylated RNA isolated from chicken cells and mouse cells transformed with recombinant copies of the chicken TK gene substantiated the map of the TK transcription unit outlined in Figure 2. As shown in Figure 4, lane 1, mouse L cells transformed with plasmid containing the 3.0 kb Hind 3 fragment produced a TK mRNA that was only 1.6 kb in length. This was about 500 bases shorter than the 2.1 kb TK mRNA produced by chicken embryo fibroblasts (lane 4). On the other hand, mouse L cells transformed with plasmid containing an additional 1.3 kb of downstream genomic sequence (the 4.3 Hind 3/Eco R1 fragment) produced a TK mRNA 2.1 kb in length (lane 2). These results reinforce our conclusion that the 4.3 kb Hind 3/R1 fragment contains the entire TK transcription unit, whereas the 3.0 kb Hind 3 fragment is missing about 500 nucleotides from the 3' terminus of the gene. Henceforth we refer to the 4.3 kb Hind 3/Eco R1 fragment as a full length chicken TK gene, abbreviated "cfl".

If the 3.0 kb Hind 3 fragment lacks the native polyadenylation signal, how are mouse cells efficiently transformed with this construct and able to produce a discreet polyadenylated TK mRNA? Functionality of the 3.0 kb Hind 3 fragment is probably due to use of a pseudo polyadenylation signal. As originally noted by Kwok et al (9) an AATTAAA heptanucleotide, which bears close resemblance to the AATAAA consensus for polyadenylation (24), is located 22 bp upstream from the 3' Hind 3 site. Use of this cryptic polyadenylation signal would predict the observed mRNA length of 1.6 kb in cells transformed with the 3.0 kb Hind 3 fragment. A low level of utilization of this cryptic polyadenylation signal was evident even in L cells transformed with the full length TK gene (note the minor signal at 1.6 kb in lane 2). Although this AATTAAA heptanucleotide was capable of serving a polyadenylation function in transformed mouse cells, the absence of even a minor 1.6 kb TK mRNA signal in chicken embryo fibroblasts (lane 4) indicates that it is not used to terminate chicken TK transcripts in homologous cells.

Table 1. TK Enzyme Activity Regulation in Myoblast Transformants^a

Genotype ^b	Clone	TK Activity Level		
		Proliferative	Committed	%Remaining
		(picomoles TMP/min-μg DNA)		
VA5'-148/tx	1E	2.82	2.88	102
	2B	2.63	1.99	76
	2C	6.07	7.40	122
cfl	1A	1.09	0.10	9
	1B	0.98	0.02	2
	1C	0.54	0.04	7
	11A	2.15	0.21	10
	11B	1.58	0.10	6
	11E	1.78	0.19	11
	11F	2.95	0.11	4
cΔ3'734/tx	1A	6.11	0.61	10
	1B	10.53	3.28	31
	1C	28.74	5.60	19
cΔi7/tx	1A	38.80	2.23	6
	5B	59.30	12.50	21
	5E	134.60	80.90	30
	5H	50.60	3.18	6
	6A	3.65	0.82	22
	6B	12.50	0.00	0
	6C	5.48	0.00	0
	6D	1.70	0.32	19
	6E	8.27	0.48	6
	6G	5.66	0.62	11

^aProliferating and committed muscle cell transformants were prepared as described in Methods and Material using a 24 hour incubation in mitogen depleted medium to induce differentiation. TK enzyme activity and DNA were determined as described previously (11). Values shown are the average of two determinations done on separate cultures.

^bGenotype refers to TK-containing plasmid used to derive transformants: vA5'-148 = HSV-TK construct with 5' terminus 148 bp upstream from cap site, a synthetic Bam H1 site immediately downstream from the translation stop codon and "tx" at its 3' deleted terminus (tx is a 379 bp 3' fragment of the HSV-TK gene carrying the viral polyadenylation signals (11)); cfl = chicken TK full length (4.3 kb Hind 3/Eco R1 fragment shown in Figure 2); cA3'734/tx = 3' deletion mutant of chicken TK ending 24 bp downstream from UGA stop codon and containing tx at its 3' terminus (11); cA17/tx = chicken TK gene lacking intron 7 (the alternatively spliced 3' nontranslated intron) and containing tx attached to a Pvu 2 site 250 bp downstream from the intron acceptor site.

Neither chicken embryo fibroblasts nor mouse cells transformed with the full length TK gene produced a TK mRNA species with the 1.3 kb size predicted for the testes-type splicing pathway. We were able to force production of a

testes-type mRNA by transforming mouse cells with a chicken TK gene from which the 3' nontranslated intron (intron 7) was removed. An intron 7 deletion mutant was constructed by replacing genomic sequence downstream from the Bgl I 2 site shown in Figure 2 with corresponding cDNA sequences from the testes cDNA clone. Since the testes cDNA fragment lacked a polyadenylation signal, a 379 bp HSV TK fragment, "tx", containing the viral polyadenylation signal (11) was attached to a restriction site near the 3' end of the testes cDNA sequence. The intron 7 deletion mutant was fully capable of transforming L cells and transformants produced large amounts of a polyadenylated TK mRNA of the size predicted for an alternatively spliced message (Figure 4, lane 3). Deletion of the 7th intron evidently did not adversely affect functionality of the TK gene or resulting mRNA. In fact the intron 7 deletion mutant consistently gave higher transformation efficiencies than the full length chicken TK gene.

We next investigated the possibility that sequences within the 3' nontranslated intron might cause the large TK mRNA to assume a configuration targeting the message for degradation in non-replicating cells and that alternative splicing removed this target. To ask whether removal of the intronic sequences resulted in escape from growth dependence, the pattern of TK enzyme activity was determined in muscle cells transformed with the intron 7 deletion mutant. As a positive and negative control for the non-regulated phenotype, the pattern of TK expression in myoblasts transformed with an HSV TK gene or a full length chicken TK gene was also determined. The results of these analyses are shown in Table 1. As previously reported (2), muscle cells transformed with the HSV TK gene (vΔ5'-148/tx) continued to express high levels of TK activity after differentiating. In contrast, muscle cells transformed with the full length chicken TK gene (cfl) or a 3' deletion mutant of the chicken TK gene containing a HSV-TK polyadenylation signal (cΔ3'734/tx) (11) both exhibited a significant loss in TK activity after differentiating. Significantly, muscle cells transformed with a chicken TK gene lacking the 3' nontranslated intron (cΔ17/tx) also showed a regulated pattern of expression. Removal of intron 7 did not result in escape from growth-dependent expression.

DISCUSSION

The experiments described above demonstrate the existence of two different-sized mRNAs encoding chicken thymidine kinase. The larger mRNA, 2100 bases in length, was the most abundant species in all cell and tissue

types examined. The smaller mRNA, 1300 bases in length, was detected only in heart and testes tissue. Southern blot analyses have established that TK is encoded by a single genetic locus (4), thus both TK mRNA species must have been generated from a common gene. Sequence analysis of cDNA clones established that the smaller message differed from the larger message in that an 863 base intron was spliced from the 3' nontranslated region of the transcript. The larger mRNA was not a nuclear processing intermediate, as it was the most abundant form of the message in both total and cytoplasmic poly A⁺ RNA preparations (data not shown).

Alternative processing of RNA transcripts has been described in several systems. In some cases functional explanations for, or consequences of, alternative splicing have been posited. For example, in many viruses, transcripts initiated from a common promoter are spliced to different reading frames on the viral genome, thus generating different proteins. The opposite occurs in the case of mouse α -amylase or *Drosophila* alcohol dehydrogenase, where transcripts initiated from different tissue- or stage-specific promoters are attached to a common protein-coding region (25,26). Alternative processing events affecting the use of different polyadenylation signals and internal protein-coding exons have also been reported (27). In the case of chicken TK, the physiological significance of alternative splicing remains obscure. Both TK mRNA species bear the same protein-coding region and 3' and 5' termini. Furthermore, the two tissues capable of producing detectable quantities of the small message, embryonic heart and adult testes, share no obvious characteristics. Transformation studies suggest that the small mRNA may be more efficiently synthesized or translated than the full length mRNA. A modified chicken TK gene lacking the alternatively-spliced 3' intron (intron 7), which is forced to produce only the small TK mRNA, consistently gave higher transformation efficiencies than the full length TK gene (data not shown). Furthermore, mouse cells transformed with the intron 7 deletion mutant generally expressed higher TK enzyme levels than mouse cells transformed with the full length gene (Table 1). These determinations are being repeated with an intron 7 deletion mutant utilizing its normal polyadenylation signal rather than a heterologous HSV-TK polyadenylation signal. If the intron 7 deletion mutant again functions more efficiently, it would indicate that sequences in the alternatively-spliced 3' nontranslated region negatively affect the synthesis, stability or translational efficiency of TK mRNA.

The possibility that alternative splicing might allow a low percentage of

TK-encoding transcripts to persist in postreplicative tissues was investigated. In view of evidence that replication-dependent expression of TK mRNA is post-transcriptionally regulated (4), a likely location for cis-acting regulatory information would be regions of the message outside of the evolutionarily constrained protein coding region. In fact, changes in stability of mouse H4 and H3 histone mRNAs during cell cycle arrest have recently been shown to be dependent on sequences within the 3' and 5' nontranslated regions of the respective messages (28, 29). By analogy with the mouse histone system, sequences within the long 3' nontranslated region of the large TK mRNA might target the transcript for degradation in non-replicating cells. Removal of most of the 3' nontranslated sequences during alternative splicing might allow the small TK mRNA to persist in postreplicative cells and supply the low levels of TK activity possibly required for basal nucleotide metabolism. Two lines of evidence suggested that this was not the case. First, comparison of TK mRNA levels in rapidly growing embryonic heart tissue and less rapidly growing hatching heart tissue indicated that both the large and small TK messages were dependent on the replicative state of the cell. Second, muscle cells transformed with a TK gene from which the 863 base nontranslated intron was removed in vitro exhibited normal down-regulation of TK activity during differentiation. These results are consistent with earlier studies showing that replacement of the chicken TK 3' non-translated region with the herpesvirus TK 3' region does not lead to loss of a regulated phenotype (2). At present, all data are consistent with our earlier conclusion that the cis-acting regulatory information responsible for replication-dependent expression of TK is located between the translation initiation and termination codons.

Acknowledgments -- This study was supported by a NIH research grant (GM34432) and New Investigator Research Award (AG05388) to GM. We are grateful to Steven McKnight for his intellectual and technical contributions to the study. We acknowledge technical assistance provided by Michael Gross, Mark Kainz and Bob Kingsbury.

REFERENCES

1. Lewis, J.A., Shimizu, K. and Zipser, D. (1983) Mol. Cell. Biol. 3, 1815-1823.
2. Merrill, G.F., Hauschka, S.D. and McKnight, S.L. (1984) Molec. Cell. Biol. 4, 1777-1784.
3. Thompson, C.B., Challoner, P.B., Neiman, P.E. and Groudine, M. (1985) Nature 314, 363-366.

4. Groudine, M. and Casimir, C. (1984) *Nuc. Acids Res.* 12, 1427-1446.
5. Leys, E.J. and Kellems, R.E. (1981) *Mol. Cell. Biol.* 1, 961-971.
6. Leys, E.J., Grouse, G.F. and Kellems, R.E. (1984) *J. Cell. Biol.* 190, 180-187.
7. Farnham, P.J. and Schimke, R.T. (1985) *J. Biol. Chem.* 260, 7675-7680.
8. Santiago, C., Collins, M. and Johnson, L.F. (1984) *J. Cell. Physiol.* 118, 79-86.
9. Kwoh, J., Zipser, D. and Wigler, M. (1983) *J. Mol. Appl. Gen.* 2, 191-200.
10. Bandyopadhyay, P.K. and Temin, H.M. (1984) *Mol. Cell. Biol.* 4, 749-754.
11. Merrill, G.F., Harland, R.M., Groudine, M. and McKnight, S.L. (1984) *Mol. Cell. Biol.* 4, 1769-1776.
12. Glisin, V., Crkvenjakow, R. and Byus, C. (1974) *Biochemistry* 13, 2633.
13. Aviv, H. and Leder, P. (1972) *Proc. Natl. Acad. Sci.* 69, 1408-1413.
14. Lehrach, H., Diamond, D., Wozney, J.M. and Boedtker, H. (1977) *Biochemistry* 16, 4743-4750.
15. Maxam, A. and Gilbert, W. (1980) *Methods Enzymol.* 65, 499-560.
16. Perucho, M., Hanahan, D., Lipsich, L. and Wigler, M. (1980) *Nature* 285, 207-211.
17. Henikoff, S. (1984) *Gene* 28, 351-359.
18. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci.* 74, 5463-5467.
19. Kit, S., Dubbs, D.R., Piekarsky, L.J. and Hsu, T.C. (1963) *Exp. Cell. Res.* 31, 297-312.
20. Graham, F. and van der Eb, A. (1973) *Virology* 52, 456-467.
21. Corsaro, C.M. and Pearson, M.L. (1981) *Somatic Cell Gen.* 7, 603-616.
22. Merrill, G.F., Witter, E. and Hauschka, S.D. (1980) *Exp. Cell Res.* 129, 191-199.
23. Linkhart, T., Clegg, C. and Hauschka, S.D. (1980) *J. Supramol. Struct.* 14, 483-498.
24. Proudfoot, N.J. and Brownlee, G.G. (1977) *Nature* 263, 211-214.
25. Hagenbuchle, O., Bovey, R. and Young, R. (1980) *Cell* 21, 179-187.
26. Benyajati, C., Spoerel, N., Haymerle, H. and Ashburner, M. (1983) *Cell* 33, 125-133.
27. Nishikura, K. and Vuocolo, G.A. (1984) *EMBO J.* 3, 689-699.
28. Luscher, B., Stauber, C., Schindler, R. and Schumperli, D. (1985) *Proc. Natl. Acad. Sci.* 82, 4389-4393.
29. Morris, T., Marashi, F., Weber, L., Hickey, E., Greenspan, D., Bonner, J., Stein, J., and Stein, G. (1986) *Proc. Natl. Acad. Sci.* 83, 981-985.